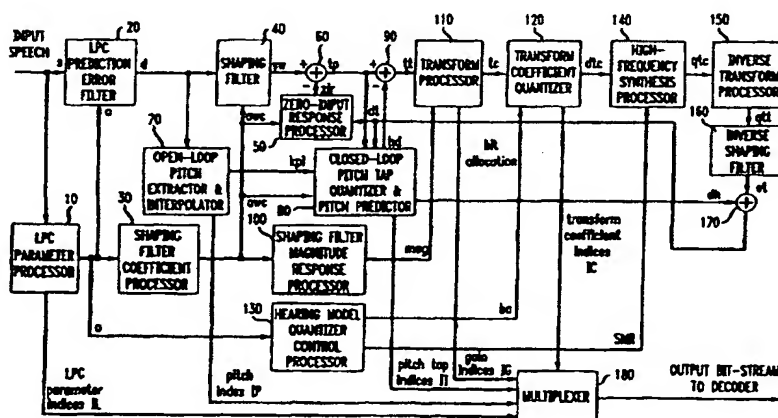




INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G10L 3/02, 9/00	A1	(11) International Publication Number: WO 97/31367 (43) International Publication Date: 28 August 1997 (28.08.97)
(21) International Application Number: PCT/US97/02898 (22) International Filing Date: 26 February 1997 (26.02.97) (30) Priority Data: 60/012,296 26 February 1996 (26.02.96) US (71) Applicant (for all designated States except US): AT & T CORP. [US/US]; 32 Avenue of the Americas, New York, NY 10013-2412 (US). (72) Inventor; and (75) Inventor/Applicant (for US only): CHEN, Juin-Hwey [CN/US]; 68 Longfield Drive, Neshanic Station, NJ 08853 (US). (74) Agent: RESTAINO, Thomas, A.; AT & T Corp., 200 Laurel Avenue, Middletown, NJ 07748 (US).		(81) Designated States: CA, JP, MX, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published With international search report.

(54) Title: MULTI-STAGE SPEECH CODER WITH TRANSFORM CODING OF PREDICTION RESIDUAL SIGNALS WITH QUANTIZATION BY AUDITORY MODELS



(57) Abstract

A speech compression system called "Transform Predictive Coding" or TPC, provides encoding for 7 kHz band speech at 16 kHz sampling at a target bit-rate of 16 or 32 kb/s one or two bits per sample. The system uses short and long term prediction to remove redundancy. The prediction residual is transformed and coded in the frequency domain as shown on the figure by (110) after accepting time domain data from (60) and parameter input from (100), which corrects the spectrum for auditory perception. The TPC coder uses only open-loop quantization as shown by (70) and therefore has low complexity. The speech quality is transparent at 32 kb/s, is very good at 24 kb/s, and is acceptable at 16 kb/s.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgyzstan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	KZ	Kazakhstan	SG	Singapore
CH	Switzerland	LJ	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovakia
CM	Cameroon	LR	Liberia	SN	Senegal
CN	China	LT	Lithuania	SZ	Swaziland
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	LV	Latvia	TG	Togo
DE	Germany	MC	Monaco	TJ	Tajikistan
DK	Denmark	MD	Republic of Moldova	TT	Trinidad and Tobago
EE	Estonia	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	UG	Uganda
FI	Finland	MN	Mongolia	US	United States of America
FR	France	MR	Mauritania	UZ	Uzbekistan
GA	Gabon			VN	Viet Nam

MULTI-STAGE SPEECH CODER WITH TRANSFORM CODING OF PREDICTION RESIDUAL SIGNALS WITH QUANTIZATION BY AUDITORY MODELS

Field of the Invention

- 5 The present invention relates to the compression (coding) of audio signals, for example, speech signals, using a predictive coding system.

Background of the Invention

- As taught in the literature of signal compression, speech and music
10 waveforms are coded by very different coding techniques. *Speech coding*, such as telephone-bandwidth (3.4 kHz) speech coding at or below 16 kb/s, has been dominated by time-domain predictive coders. These coders use speech *production* models to *predict* speech waveforms to be coded. Predicted waveforms are then subtracted from the actual (original) waveforms
15 (to be coded) to reduce redundancy in the original signal. Reduction in signal redundancy provides coding gain. Examples of such *predictive* speech coders include Adaptive Predictive Coding, Multi-Pulse Linear Predictive Coding, and Code-Excited Linear Prediction (CELP) Coding, all well known in the art of speech signal compression.
- 20 On the other hand, wideband (0 - 20 kHz) music coding at or above 64 kb/s has been dominated by frequency-domain transform or sub-band coders. These music coders are fundamentally very different from the speech coders discussed above. This difference is due to the fact that the sources of music, unlike those of speech, are too varied to allow ready prediction.
- 25 Consequently, models of music sources are generally not used in music coding. Instead, music coders use elaborate human hearing models to code only those parts of the signal that are perceptually relevant. That is, unlike speech coders which commonly use speech production models, music coders employ hearing – *sound reception* – models to obtain coding gain.
- 30 In music coders, hearing models are used to determine a noise masking capability of the music to be coded. The term “noise masking capability” refers to how much quantization noise can be introduced into a

music signal without a listener noticing the noise. This noise masking capability is then used to set quantizer resolution (e.g., quantizer stepsize). Generally, the more "tonelike" music is, the poorer the music will be at masking quantization noise and, therefore, the smaller the required stepsize
5 will be, and *vice versa*. Smaller stepsizes correspond to smaller coding gains, and *vice versa*. Examples of such music coders include AT&T's Perceptual Audio Coder (PAC) and the ISO MPEG audio coding standard.

In between telephone-bandwidth speech coding and wideband music coding, there lies *wideband speech coding*, where the speech signal is
10 sampled at 16 kHz and has a bandwidth of 7 kHz. The advantage of 7 kHz wideband speech is that the resulting speech quality is much better than telephone-bandwidth speech, and yet it requires a much lower bit-rate to code than a 20 kHz audio signal. Among those previously proposed wideband speech coders, some use time-domain predictive coding, some use
15 frequency-domain transform or sub-band coding, and some use a mixture of time-domain and frequency-domain techniques.

The inclusion of perceptual criteria in predictive speech coding, wideband or otherwise, has been limited to the use of a perceptual weighting filter in the context of selecting the best synthesized speech signal from
20 among a plurality of candidate synthesized speech signals. See, e.g., U.S. Patent No. Re. 32,580 to Atal *et al*. Such filters accomplish a type of noise shaping which is useful reducing noise in the coding process. One known coder attempts to improve upon this technique by employing a perceptual model in the formation of that perceptual weighting filter.

Summary of the Invention

The efforts described above notwithstanding, none of the known speech or audio coders utilizes both a *speech production model* for signal prediction purposes and a *hearing model* to set quantizer resolution according to an analysis of signal noise masking capability.

The present invention, on the other hand, combines a predictive coding system with a quantization process which quantizes a signal based on a noise masking signal determined with a model of human auditory sensitivity to noise. The output of the predictive coding system is thus quantized with a quantizer having a resolution (e.g., stepsize in a uniform scalar quantizer, or the number of bits used to identify vectors in a vector quantizer) which is a function of a noise masking signal determined in accordance with a audio perceptual model.

According to the invention, a signal is generated which represents an estimate (or prediction) of a signal representing speech information. The term "original signal representing speech information" is broad enough to refer not only to speech itself, but also to speech signal derivatives commonly found in speech coding systems (such as linear prediction and pitch prediction residual signals). The estimate signal is then compared to the original signal to form a signal representing the difference between said compared signals. This signal representing the difference between the compared signals is then quantized in accordance with a perceptual noise masking signal which is generated by a model of human audio perception.

An illustrative embodiment of the present invention, referred to as "Transform Predictive Coding", or TPC, encodes 7 kHz wideband speech at a target bit-rate of 16 to 32 kb/s. As its name implies, TPC combines transform coding and predictive coding techniques in a single coder. More specifically, the coder uses linear prediction to remove the redundancy from the input speech waveform and then use transform coding techniques to encode the resulting prediction residual. The transformed prediction residual is quantized based on knowledge in human auditory perception, expressed in terms of a

auditory perceptual model, to encode what is audible and discard what is inaudible.

One important feature of the illustrative embodiment concerns the way in which perceptual noise masking capability (e.g., the perceptual threshold of "just noticeable distortion") of the signal is determined and subsequent bit allocation is performed. Rather than determining a perceptual threshold using the unquantized input signal, as is done in conventional music coders, the noise masking threshold and bit allocation of the embodiment are determined based on the frequency response of a quantized synthesis filter – in the embodiment, a quantized LPC synthesis filter. This feature provides an advantage to the system of not having to communicate bit allocation signals, from the encoder to the decoder, in order for the decoder to replicate the perceptual threshold and bit allocation processing needed for decoding the received coded wideband speech information. Instead, synthesis filter coefficients, which are being communicated for other purposes, are exploited to save bit rate.

Another important feature of the illustrative embodiment concerns how the TPC coder allocates bits among coder frequencies and how the decoder generates a quantized output signal based on the allocated bits. In certain circumstances, the TPC coder allocates bits only to a portion of the audio band (for example, bits may be allocated to coefficients between 0 and 4 kHz, only). No bits are allocated to represent coefficients between 4 kHz and 7 kHz and, thus, the decoder gets no coefficients in this frequency range. Such a circumstance occurs when, for example, the TPC coder has to operate at very low bit rates, e.g., 16 kb/s. Despite having no bits representing the coded signal in the 4 kHz and 7 kHz frequency range, the decoder must still synthesize a signal in this range if it is to provide a wideband response. According to this feature of the embodiment, the decoder generates - that is, synthesizes - coefficient signals in this range of frequencies based on other available information - a ratio of an estimate of the signal spectrum (obtained from LPC parameters) to a noise masking threshold at frequencies in the range. Phase values for the coefficients are selected at random. By virtue of

this technique, the decoder can provide a wideband response without the need to transmit speech signal coefficients for the entire band.

The potential applications of a wideband speech coder include ISDN video-conferencing or audio-conferencing, multimedia audio, "hi-fi" telephony, and simultaneous voice and data (SVD) over dial-up lines using modems at 28.8 kb/s or higher.

Brief Description of the Drawings

Figure 1 presents an illustrative coder embodiment of the present invention.

Figure 2 presents an illustrative decoder embodiment of the present invention.

Figure 3 presents a detailed block diagram of the LPC parameter processor of Figure 1.

Detailed Description

A. Overview of the Illustrative Embodiments

For clarity of explanation, the illustrative embodiment of the present invention is presented as comprising individual functional blocks (including functional blocks labeled as "processors"). The functions these blocks represent may be provided through the use of either shared or dedicated hardware, including, but not limited to, hardware capable of executing software. For example, the functions of processors presented in Figures 1 to 4 may be provided by a single shared processor. (Use of the term "processor" should not be construed to refer exclusively to hardware capable of executing software.)

Illustrative embodiments may comprise digital signal processor (DSP) hardware, such as the AT&T DSP16 or DSP32C, read-only memory (ROM) for storing software

performing the operations discussed below, and random access memory (RAM) for storing DSP results. Very large scale integration (VLSI) hardware embodiments, as well as custom VLSI circuitry in combination with a general purpose DSP circuit, may also be provided.

5 In accordance with the present invention, the sequence of digital input speech samples is partitioned into consecutive 20 ms blocks called *frames*, and each frame is further subdivided into 5 equal *subframes* of 4 ms each. Assuming a sampling rate of 16 kHz, as is common for wideband speech signals, this corresponds to a frame size of 320 samples and a subframe size
10 of 64 samples. The TPC speech coder buffers and processes the input speech signal frame-by-frame, and within each frame certain encoding operations are performed subframe-by-subframe.

Figure 1 presents an illustrative TPC speech coder embodiment of the present invention. Refer to the embodiment shown in Figure 1. Once every
15 20 ms frame, the LPC parameter processor 10 derives the Line Spectral Pair (LSP) parameters from the input speech signal s , quantizes such LSP parameters, interpolates them for each 4 ms subframe, and then converts to the LPC predictor coefficient array a for each subframe. Short-term redundancy is removed from the input speech signal, s , by the LPC prediction
20 error filter 20. The resulting LPC prediction residual signal, d , still has some long-term redundancy due to the pitch periodicity in voiced speech. The shaping filter coefficient processor 30 derives the shaping filter coefficients awc from quantized LPC filter coefficients a . The shaping filter 40 filters the LPC prediction residual signal d to produce a perceptually weighted speech
25 signal sw . The zero-input response processor 50 calculates the zero-input response, zir , of the shaping filter. The subtracting unit 60 then subtracts zir from sw to obtain tp , the target signal for pitch prediction.

The open-loop pitch extractor and interpolator 70 uses the LPC prediction residual d to extract a pitch period for each 20 ms frame, and then
30 calculates the interpolated pitch period kpi for each 4 ms sub-frame. The closed-loop pitch tap quantizer and pitch predictor 80 uses this interpolated pitch period kpi to select a set of 3 pitch predictor taps from a codebook of

candidate sets of pitch taps. The selection is done such that when the previously quantized LPC residual signal dt is filtered by the corresponding 3-tap pitch synthesis filter and then by a shaping filter with zero initial memory, the output signal hd is closest to the target signal tp in a mean-square error (MSE) sense. The subtracting unit 90 subtracts hd from tp to obtain tt , the target signal for transform coding.

The shaping filter magnitude response processor 100 calculates the signal mag , the magnitude of the frequency response of the shaping filter. The transform processor 110 performs a linear transform, such as Fast Fourier Transform (FFT), on the signal tt . Then, it normalizes the transform coefficients using mag and the quantized versions of gain values which is calculated over three different frequency bands. The result is the normalized transform coefficient signal tc . The transform coefficient quantizer 120 then quantizes the signal tc using the adaptive bit allocation signal ba , which is determined by the hearing model quantizer control processor 130 according to the time-varying perceptual importance of transform coefficients at different frequencies.

At a lower bit-rate, such as 16 kb/s, processor 130 only allocates bits to the lower half of the frequency band (0 to 4 kHz). In this case, the high-frequency synthesis processor 140 synthesizes the transform coefficients in the high-frequency band (4 to 8 kHz), and combine them with the quantized low-frequency transform coefficient signal d_{tc} to produce the final quantized full-band transform coefficient signal q_{tc} . At a higher bit-rate, such as 24 or 32 kb/s, each transform coefficient in the entire frequency band is allowed to receive bits in the adaptive bit allocation process, although coefficients may eventually receive no bits at all due to the scarcity of the available bits. In this case, the high-frequency synthesis processor 140 simply detects those frequencies in the 4 to 8 kHz band that receive no bits, and fills in such "spectral holes" with low-level noise to avoid a type of "swirling" distortion typically found in adaptive transform coders.

The inverse transform processor 150 takes the quantized transform coefficient signal q_{tc} , and applies a linear transform which is the inverse

operation of the linear transform employed in the transform processor 110 (an inverse FFT in our particular illustrative embodiment here). This results in a time-domain signal q_{tt} , which is the quantized version of tt , the target signal for transform coding. The inverse shaping filter 160 then filters q_{tt} to obtain
5 the quantized excitation signal ef . The adder 170 adds ef to the signal dh (which is the pitch-predicted version of the LPC prediction residual d) produced by the pitch predictor inside block 80. The resulting signal dt is the quantized version of the LPC prediction residual d . It is used to update the filter memory of the shaping filter inside the zero-input response processor 50
10 and the memory of the pitch predictor inside block 80. This completes the signal loop.

Codebook indices representing the LPC predictor parameters (IL), the pitch predictor parameters (IP and IT), the transform gain levels (IG), and the quantized transform coefficients (IC) are multiplexed into a bit stream by the
15 multiplexer 180 and transmitted over a channel to a decoder. The channel may comprise any suitable communication channel, including wireless channels, computer and data networks, telephone networks; and may include or consist of memory, such as, solid state memories (for example, semiconductor memory), optical memory systems (such as CD-ROM),
20 magnetic memories (for example, disk memory), etc.

Figure 2 presents an illustrative TPC speech decoder embodiment of the present invention. The demultiplexer 200 separates the codebook indices IL , IP , IT , IG , and IC . The pitch decoder and interpolator 205 decodes IP and calculates the interpolated pitch period k_{pi} . The pitch tap decoder and pitch
25 predictor 210 decodes IT to obtain the pitch predictor taps array b , and it also calculates the signal dh , or the pitch-predicted version of the LPC prediction residual d . The LPC parameter decoder and interpolator 215 decodes IL and then calculates the interpolated LPC filter coefficient array a . Blocks 220 through 255 perform exactly the same operations as their counterparts in
30 Figure 1 to produce the quantized LPC residual signal dt . The long-term postfilter 260 enhances the pitch periodicity in dt and produces a filtered version fdt as its output. This signal is passed through the LPC synthesis

filter 265, and the resulting signal st is further filtered by the short-term postfilter 270, which produces a final filtered output speech signal fst .

To keep the complexity low, open-loop quantization is employed by the TPC as much as possible. Open-loop quantization means the quantizer
5 attempts to minimize the difference between the unquantized parameter and its quantized version, without regard to the effects on the output speech quality. This is in contrast to, for example, CELP coders, where the pitch predictor, the gain, and the excitation are usually close-loop quantized. In closed-loop quantization of a coder parameter, the quantizer codebook
10 search attempts to minimize the distortion in the final reconstructed output speech. Naturally, this generally leads to a better output speech quality, but at the price of a higher codebook search complexity.

In the present invention, the TPC coder uses closed-loop quantization only for the 3 pitch predictor taps. The quantization operations leading to the
15 quantized excitation signal ef is basically similar to open-loop quantization, but the effects on the output speech is close to that of closed-loop quantization. This approach is similar in spirit to the approach used in the TCX coder by Lefebvre et. al., "High Quality Coding of Wideband Audio Signals Using Transform Coded Excitation (TCX)", Proc. IEEE International
20 Conf. Acoustics, Speech, Signal Processing, 1994, pp. I-193 to I-196, although there are also important differences. For example, the features of the current invention that are not in the TCX coder include normalization of the transform coefficients by a shaping filter magnitude response, adaptive bit allocation controlled by a hearing model, and the high-frequency synthesis
25 and noise fill-in procedures.

B. An Illustrative Coder Embodiment

1. The LPC Analysis and Prediction

A detailed block diagram of LPC parameter processor 10 is presented in Figure 3. Processor 10 comprises a windowing and autocorrelation
30 processor 310; a spectral smoothing and white noise correction processor

315; a Levinson-Durbin recursion processor 320; a bandwidth expansion processor 325; an LPC to LSP conversion processor 330; and LPC power spectrum processor 335; an LSP quantizer 340; an LSP sorting processor 345; an LSP interpolation processor 350; and an LSP to LPC conversion
5 processor 355.

Windowing and autocorrelation processor 310 begins the process of LPC coefficient generation. Processor 310 generates autocorrelation coefficients, r , in conventional fashion, once every 20 ms from which LPC coefficients are subsequently computed, as discussed below. See Rabiner,
10 L. R. *et al.*, Digital Processing of Speech Signals, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1978 (Rabiner *et al.*). The LPC frame size is 20 ms (or 320 speech samples at 16 kHz sampling rate). Each 20 ms frame is further divided into 5 subframes, each 4 ms (or 64 samples) long. LPC analysis processor uses a 24 ms Hamming window which is centered at the
15 last 4 ms subframe of the current frame, in conventional fashion.

To alleviate potential ill-conditioning, certain conventional signal conditioning techniques are employed. A spectral smoothing technique (SST) and a white noise correction technique are applied by spectral smoothing and white noise correction processor 315 before LPC analysis. The SST, well-
20 known in the art (Tohkura, Y. *et al.*, "Spectral Smoothing Technique in PARCOR Speech Analysis-Synthesis," IEEE Trans. Acoust., Speech, Signal Processing, ASSP-26:587-596, December 1978 (Tohkura *et al.*)) involves multiplying an calculated autocorrelation coefficient array (from processor 310) by a Gaussian window whose Fourier transform corresponds to a
25 probability density function (pdf) of a Gaussian distribution with a standard deviation of 40 Hz. The white noise correction, also conventional (Chen, J.-H., "A Robust Low-Delay CELP Speech Coder at 16 kbit/s, Proc. IEEE Global Comm. Conf., pp. 1237-1241, Dallas, TX, November 1989.), increases the zero-lag autocorrelation coefficient (*i.e.*, the energy term) by 0.001%.

30 The coefficients generated by processor 315 are then provided to Levinson-Durbin recursion processor 320, which generates 16 LPC

coefficients, α_i for $i=1,2,\dots,16$ (the order of the LPC prediction error filter 20 is 16) in conventional fashion.

Bandwidth expansion processor 325 multiplies each α_i by a factor g^i , where $g=0.994$, for further signal conditioning. This corresponds to a bandwidth
5 expansion of 30 Hz. (Tohkura et al.).

After such a bandwidth expansion, the LPC predictor coefficients are converted to the Line Spectral Pair (LSP) coefficients by LPC to LSP conversion processor 330 in conventional fashion. See Soong, F. K. et al., "Line Spectrum Pair (LSP) and Speech Data Compression," Proc. IEEE Int.
10 Conf. Acoust., Speech, Signal Processing, pp. 1.10.1-1.10.4, March 1984 (Soong et al.), which is incorporated by reference as if set forth fully herein.

Vector quantization (VQ) is then provided by LSP quantizer 340 to quantize the resulting LSP coefficients. The specific VQ technique employed by processor 240 is similar to the split VQ proposed in Paliwal, K. K. et al.,
15 "Efficient Vector Quantization of LPC Parameters at 24 bits/frame," Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, pp. 661-664, Toronto, Canada, May 1991 (Paliwal et al.), which is incorporated by reference as if set forth fully herein. The 16-dimensional LSP vector is split into 7 smaller sub-vectors having the dimensions of 2, 2, 2, 2, 2, 3, 3, counting from the low-
20 frequency end. Each of the 7 sub-vectors are quantized to 7 bits (i.e., using a VQ codebook of 128 codevectors). Thus, there are seven codebook indices, $IL(1) - IL(7)$, each index being seven bits in length, for a total of 49 bits per frame used in LPC parameter quantization. These 49 bits are provided to the multiplexer 180 for transmission to the decoder as side information.

25 Processor 340 performs its search through the VQ codebook using a conventional weighted mean-square error (WMSE) distortion measure, as described in Paliwal et al. The LPC power spectrum processor 335 is used to calculate the weights in this WMSE distortion measure. The codebook used in processor 340 is designed with conventional codebook generation
30 techniques well-known in the art. A conventional MSE distortion measure

can also be used instead of the WMSE measure to reduce the coder's complexity without significant degradation in the output speech quality.

Normally LSP coefficients monotonically increase. However, quantization may result in a disruption of this order. This disruption results in
5 an unstable LPC synthesis filter in the decoder. To avoid this problem, the LSP sorting processor 345 sorts the quantized LSP coefficients to restore the monotonically increasing order and ensure stability.

The quantized LSP coefficients are used in the last subframe of the current frame. Linear interpolation between these LSP coefficients and those
10 from the last subframe of the previous frame is performed to provide LSP coefficients for the first four subframes by LSP interpolation processor 350, as is conventional. The interpolated and quantized LSP coefficients are then converted back to the LPC predictor coefficients for use in each subframe by LSP to LPC conversion processor 355 in conventional fashion. This is done
15 in both the encoder and the decoder. The LSP interpolation is important in maintaining the smooth reproduction of the output speech. The LSP interpolation allows the LPC predictor coefficients to be updated once a subframe (4 ms) in a smooth fashion. The resulting LPC predictor coefficient array a are used in the LPC prediction error filter 20 to predict the coder's
20 input signal. The difference between the input signal and its predicted version is the LPC prediction residual, d .

2. Shaping Filter

The shaping filter coefficient processor 30 computes the first three autocorrelation coefficients of the LPC predictor coefficient array a , then uses
25 the Levinson-Durbin recursion to solve for the coefficients $c_j, j = 0,1,2$ for the corresponding optimal second-order all-pole predictor. These predictor coefficients are then bandwidth-expanded by a factor of 0.7 (i.e. the j -th coefficient c_j is replaced by $c_j(0.7)^j$). Next, processor 30 also performs bandwidth expansion of the 16th-order all-pole LPC predictor coefficient array
30 a , but this time by a factor of 0.8. Cascading these two bandwidth-expanded

all-pole filters (2nd-order and 16th-order) gives us the desired 18th-order shaping filter 40. The shaping filter coefficient array *awc* is calculated by convolving the two bandwidth-expanded coefficient arrays (2nd-order and 16th-order) mentioned above to get a direct-form 18th-order filter.

- 5 When the shaping filter 40 are cascaded with the LPC prediction error filter, as is shown in Figure 1, the two filters effectively form a perceptual weighting filter whose frequency response is roughly the inverse of the desired coding noise spectrum. Thus, the output of the shaping filter 40 is called the perceptually weighted speech signal *sw*.
- 10 The zero-input response processor 50 has a shaping filter in it. At the beginning of each 4 ms subframe, it performs shaping filtering by feeding the filter with 4 ms worth of zero input signal. In general, the corresponding output signal vector *zir* is non-zero because the filter generally has non-zero memory (except during the very first subframe after coder initialization, or
- 15 when the coder's input signal has been exactly zero since the coder starts up). Processor 60 subtracts *zir* from the weighted speech vector *sw*; the resulting signal vector *tp* is the target vector for closed-loop pitch prediction.

3. Closed-Loop Pitch Prediction

- 20 There are two kinds of parameters in pitch prediction which need to be quantized and transmitted to the decoder: the pitch period corresponding to the period of the nearly periodic waveform of voiced speech, and the three pitch predictor coefficients (taps).

25 a. Pitch Period

- The pitch period of the LPC prediction residual is determined by the open-loop pitch extractor and interpolator 70 using a modified version of the efficient two-stage search technique discussed in U.S. Patent No. 5,327,520, entitled "Method of Use of Voice Message Coder/Decoder," and incorporated
- 30 by reference as if set forth fully herein. Processor 70 first passes the LPC residual through a third-order elliptic lowpass filter to limit the bandwidth to about 700 Hz, and then performs 8:1 decimation of the lowpass filter output.

Using a pitch analysis window corresponding to the last 3 subframes of the current frame, the correlation coefficients of the decimated signal are calculated for time lags ranging from 3 to 34, which correspond to time lags of 24 to 272 samples in the undecimated signal domain. Thus, the allowable
5 range for the pitch period is 1.5 ms to 17 ms, or 59 Hz to 667 Hz in terms of the pitch frequency. This is sufficient to cover the normal pitch range of most speakers, including low-pitched males and high-pitched children.

After the correlation coefficients of the decimated signal are calculated, the first major peak of the correlation coefficients which has the lowest time
10 lag is identified. This is the first-stage search. Let the resulting time lag be t . This value t is multiplied by 8 to obtain the time lag in the undecimated signal domain. The resulting time lag, $8t$, points to the neighborhood where the true pitch period is most likely to lie. To retain the original time resolution in the undecimated signal domain, a second-stage pitch search is conducted in the
15 range of $t-4$ to $t+4$. The correlation coefficients of the original undecimated LPC residual, d , are calculated for the time lags of $t-4$ to $t+4$ (subject to the lower bound of 24 samples and upper bound of 272 samples). The time lag corresponding to the maximum correlation coefficient in this range is then identified as the final pitch period. This pitch period is encoded into 8 bits and
20 the 8-bit index IP is provided to the multiplexer 180 for transmission to the decoder as side information. Eight bits are sufficient to represent the pitch period since there are only $272-24+1=249$ possible integers that can be selected as the pitch period.

Only one such 8-bit pitch index is transmitted for each 20 ms frame.
25 Processor 70 determines the pitch period kpi for each subframe in the following way. If the difference between the extracted pitch period of the current frame and that of the last frame is greater than 20%, the extracted pitch period described above is used for every subframe in the current frame.

On the other hand, if this relative pitch change is less than 20%, then the
30 extracted pitch period is used for the last 3 subframes of the current frame, while the pitch periods of the first 2 subframes are obtained by a linear

interpolation between the extracted pitch period of the last frame and that of the current frame.

b. Pitch Predictor Taps

5 The closed-loop pitch tap quantizer and pitch predictor 80 performs the following operations subframe-by-subframe: (1) closed-loop quantization of the 3 pitch taps, (2) generation of dh , the pitch-predicted version of the LPC prediction residual d in the current subframe, and (3) generation of hd , the closest match to the target signal tp .

10 Processor 80 has an internal buffer that stores previous samples of the signal dt , which can be regarded as the quantized version of the LPC prediction residual d . For each subframe, processor 80 uses the pitch period kpi to extract three 64-dimensional vectors from the dt buffer. These three vectors, which are called x_1, x_2 , and x_3 , are respectively $kpi - 1$, kpi , and kpi
 15 $+ 1$ samples earlier than the current frame of dt . These three vectors are then separately filtered by a shaping filter (with the coefficient array awc) which has zero initial filter memory. Let's call the resulting three 64-dimensional output vectors y_1, y_2 , and y_3 . Next, processor 80 needs to search through a codebook of 64 candidate sets of 3 pitch predictor taps
 20 $b_{1j}, b_{2j}, b_{3j}, j = 1, 2, \dots, 64$, and find the optimal set b_{1k}, b_{2k}, b_{3k} which minimizes the distortion measure

$$\|tp - \sum_{i=1}^3 b_{ik} y_i\|^2.$$

This type of problem has been studied before, and an efficient search method can be found in U.S. Patent No. 5,327,520. While the details of this technique
 25 will not be presented here, the basic idea is as follows.

It can be shown that minimizing this distortion measure is equivalent to maximizing an inner product of two 9-dimensional vectors. One of these 9-dimensional vectors contains only correlation coefficients of y_1, y_2 , and y_3 . The other 9-dimensional vector contains only the product terms derived from
 30 the set of three pitch predictor taps under evaluation. Since such a vector is signal-independent and depends only on the pitch tap codevector, there are

only 64 such possible vectors (one for each pitch tap codevector), and they can be pre-computed and stored in a table—the VQ codebook. In an actual codebook search, the 9-dimensional correlation vector of y_1, y_2 , and y_3 is calculated first. Next, the inner product of the resulting vector with each of the

5 64 pre-computed and stored 9-dimensional vectors is calculated. The vector in the stored table which gives the maximum inner product is the winner, and the three quantized pitch predictor taps are derived from it. Since there are 64 vectors in the stored table, a 6-bit index, $IT(m)$ for the m -th subframe, is sufficient to represent the three quantized pitch predictor taps. Since there

10 are 5 subframes in each frame, a total of 30 bits per frame are used to represent the three pitch taps used for all subframes. These 30 bits are provided to the multiplexer 180 for transmission to the decoder as side information.

For each subframe, after the optimal set of 3 pitch taps b_{1k}, b_{2k}, b_{3k} are

15 selected by the codebook search method outlined above, the pitch-predicted version of d is calculated as

$$dh = \sum_{i=1}^3 b_{ik} x_i .$$

The output signal vector hd is calculated as

$$hd = \sum_{i=1}^3 b_{ik} y_i .$$

20 This vector hd is subtracted from the vector tp by the subtracting unit 90. The result is tt , the target vector for transform coding.

4. Transform Coding of the Target Vector

a. Shaping Filter Magnitude Response for Normalization

25 The target vector tt is encoded subframe-by-subframe by blocks 100 through 150 using a transform coding approach. The shaping filter magnitude response processor 100 calculates the signal mag in the following way. First, it takes the shaping filter coefficient array awc of the last subframe of the

30 current frame, zero-pads it to 64 samples, and then performs a 64-point FFT

on the resulting 64-dimensional vector. Then, it calculates the magnitudes of the 33 FFT coefficients which correspond to the frequency range of 0 to 8 kHz. The result vector *mag* is the magnitude response of the shaping filter for the last subframe. To save computation, the *mag* vectors for the first four
5 subframes are obtained by a linear interpolation between the *mag* vector of the last subframe of the last frame and that of the last subframe of the current frame.

b. Transform and Gain Normalization

The transform processor 110 performs several operations, as
10 described below. It first transforms the 64-dimensional vector *tt* in the current subframe by using a 64-point FFT. This transform size of 64 samples (or 4 ms) avoids the so-called "pre-echo" distortion well-known in the audio coding art. See Jayant, N. et al., "Signal Compression Based on Models of Human Perception," Proc. IEEE, pp. 1385-1422, October 1993 which is incorporated
15 by reference as if set forth fully herein. Each of the first 33 complex FFT coefficients is then divided by the corresponding element in the *mag* vector. The resulting normalized FFT coefficient vector is partitioned into 3 frequency bands: (1) the low-frequency band consisting of the first 6 normalized FFT coefficients (i.e. from 0 to 1250 Hz), (2) the mid-frequency band consisting of
20 the next 10 normalized FFT coefficients (from 1500 to 3750 Hz), and (3) the high-frequency band consisting of the remaining 17 normalized FFT coefficients (from 4000 to 8000 Hz).

The total energy in each of the 3 bands are calculated and then converted to dB value, called the *log gain* of each band. The log gain of the
25 low-frequency band is quantized using a 5-bit scalar quantizer designed using the Lloyd algorithm well known in the art. The quantized low-frequency log gain is subtracted from the log gains of the mid- and high- frequency bands. The resulting level-adjusted mid- and high-frequency log gains are concatenated to form a 2-dimensional vector, which is then quantized by a 7-
30 bit vector quantizer, with a codebook designed by the generalized Lloyd algorithm, again well-known in the art. The quantized low-frequency log gain is then added back to the quantized versions of the level-adjusted mid- and

high-frequency log gains to obtain the quantized log gains of the mid- and high-frequency bands. Next, all three quantized log gains are converted from the logarithmic (dB) domain back to the linear domain. Each of the 33 normalized FFT coefficients (normalized by *mag* as described above) is then
 5 further divided by the corresponding quantized linear gain of the frequency band where the FFT coefficient lies in. After this second stage of normalization, the result is the final normalized transform coefficient vector *tc*, which contains 33 complex numbers representing frequencies from 0 to 8000 Hz.

10 During the quantization of log gains in the *m*-th subframe, the transform processor 110 produces a 5-bit gain codebook index *IG(m,1)* for the low-frequency log gain and a 7-bit gain codebook index *IG(m,2)* for the mid- and high-frequency log gains. Therefore, the 3 log gains are encoded at a bit-rate of 12 bits per subframe, or 60 bits per frame. These 60 bits are
 15 provided to the multiplexer 180 for transmission to the decoder as side information. These 60 gain bits, along with the 49 bits for LSP, 8 bits for the pitch period, and 30 bits for the pitch taps, form the side information, which totals $49 + 8 + 30 + 60 = 147$ bits per frame.

c. The Bit Stream

20 As described above, 49 bits/frame have been allocated for encoding LPC parameters, $8 + (6 \times 5) = 38$ bits/frame have been allocated for the 3-tap pitch predictor, and $(5 + 7) \times 5 = 60$ bits/frame for the gains. Therefore, the total number of side information bits is $49 + 38 + 60 = 147$ bits per 20 ms frame, or roughly 30 bits per 4 ms subframe. Consider that the coder might be used at
 25 one of three different rates: 16, 24 and 32 kb/s. At a sampling rate of 16 kHz, these three target rates translate to 1, 1.5, and 2 bits/sample, or 64, 96, and 128 bits/subframe, respectively. With 30 bits/subframe used for side information, the numbers of bits remaining to use in encoding the main information (encoding of FFT coefficients) are 34, 66, and 98 bits/subframe
 30 for the three rates of 16, 24, and 32 kb/s, respectively.

d. Adaptive Bit Allocation

In accordance with the principles of the present invention, adaptive bit allocation is performed to assign these remaining bits to various parts of the frequency spectrum with different quantization accuracy, in order enhance the
5 *perceptual quality* of the output speech at the TPC decoder. This is done by using a model of human sensitivity to noise in audio signals. Such models are known in the art of perceptual audio coding. See, e.g., Tobias, J. V., ed., Foundations of Modern Auditory Theory, Academic Press, New York and London, 1970. See also Schroeder, M. R. *et al.*, "Optimizing Digital Speech
10 Coders by Exploiting Masking Properties of the Human Ear," J. Acoust. Soc. Amer., 66:1647-1652, December 1979 (Schroeder, *et al.*), which is hereby incorporated by reference as if fully set forth herein.

Hearing model and quantizer control processor 130 performs adaptive bit allocation and generate an output vector *ba* which tells the transform
15 coefficient quantizer 120 how many bits should be used to quantize each of the 33 normalized transform coefficients contained in *tc*. While adaptive bit allocation might be performed once every subframe, the illustrative embodiment of the present invention performs bit allocation once per frame in order to reduce computational complexity.

20 Rather than using the unquantized input signal to derive the noise masking threshold and bit allocation, as is done in conventional music coders, the noise masking threshold and bit allocation of the illustrative embodiment are determined from the frequency response of the quantized LPC synthesis filter (which is often referred to as the "LPC spectrum"). The LPC spectrum
25 can be considered an approximation of the spectral envelope of the input signal within the 24 ms LPC analysis window. The LPC spectrum is determined based on the quantized LPC coefficients. The quantized LPC coefficients are provided by the LPC parameter processor 10 to the hearing model and quantizer control processor 130, which determines the LPC
30 spectrum as follows. The quantized LPC filter coefficients *a* are first transformed by a 64-point FFT. The power of each of the first 33 FFT coefficients is determined and the reciprocal is then calculated. The result is

the LPC power spectrum which has the frequency resolution of a 64-point FFT.

After the LPC power spectrum is determined, an estimated noise masking threshold, T_M , is calculated using a modified version of the method described in U.S. Patent No. 5,314,457, which is incorporated by reference as if fully set forth herein. Processor 130 scales the 33 samples of LPC power spectrum by a frequency-dependent attenuation function empirically determined from subjective listening experiments. The attenuation function starts at 12 dB for the DC term of the LPC power spectrum, increases to about 15 dB between 700 and 800 Hz, then decreases monotonically toward high frequencies, and finally reduces to 6 dB at 8000 Hz.

Each of the 33 attenuated LPC power spectrum samples is then used to scale a "basilar membrane spreading function" derived for that particular frequency to calculate the masking threshold. A spreading function for a given frequency corresponds to the shape of the masking threshold in response to a single-tone masker signal at that frequency. Equation (5) of Schroeder, *et al.* describes such spreading functions in terms of the "bark" frequency scale, or critical-band frequency scale is incorporated by reference as if set forth fully herein. The scaling process begins with the first 33 frequencies of a 64-point FFT (i.e., 0 Hz, 250 Hz, 500 Hz, . . . , 8000 Hz) being converted to the "bark" frequency scale. Then, for each of the 33 resulting bark values, the corresponding spreading function is sampled at these 33 bark values using equation (5) of Schroeder *et al.* The 33 resulting spreading functions are stored in a table, which may be done as part of an off-line process. To calculate the estimated masking threshold, each of the 33 spreading functions is multiplied by the corresponding sample value of the attenuated LPC power spectrum, and the resulting 33 scaled spreading functions are summed together. The result is the estimated masking threshold function. It should be noted that this technique for estimating the masking threshold is not the only technique available.

To keep the complexity low, processor 130 uses a "greedy" algorithm to perform adaptive bit allocation. The technique is "greedy" in the sense that

it allocates one bit at a time to the most "needy" frequency component without regard to its potential influence on future bit allocation. At the beginning when no bit is assigned yet, the corresponding output speech will be zero, and the coding error signal is the input speech itself. Therefore, initially the LPC
 5 power spectrum is assumed to be the power spectrum of the coding noise. Then, the noise loudness at each of the 33 frequencies of a 64-point FFT is estimated using the masking threshold calculated above and a simplified version of the noise loudness calculation method in Schroeder *et al.*

The simplified noise loudness at each of the 33 frequencies is
 10 calculated as follows. First, the critical bandwidth B_i at the i -th frequency is calculated using linear interpolation of the critical bandwidth listed in table 1 of Scharf's book chapter in Tobias. The result is the approximated value of the term $d\bar{f}/dx$ in equation (3) of Schroeder *et al.* The 33 critical bandwidth values are pre-computed and stored in a table. Then, for the i -th frequency, the
 15 noise power N_i is compared with the masking threshold M_i . If $N_i \leq M_i$, the noise loudness L_i is set to zero. If $N_i > M_i$, then the noise loudness is calculated as

$$L_i = B_i ((N_i - M_i) / (1 + (S_i / N_i)^2))^{0.25}$$

where S_i is the sample value of the LPC power spectrum at the i -th frequency.

20

Once the noise loudness is calculated for all 33 frequencies, the frequency with the maximum noise loudness is identified and one bit is assigned to this frequency. The noise power at this frequency is then reduced by a factor which is empirically determined from the signal-to-noise ratio
 25 (SNR) obtained during the design of the VQ codebook for quantizing the normalized FFT coefficients. (Illustrative values for the reduction factor are between 4 and 5 dB). The noise loudness at this frequency is then updated using the reduced noise power. Next, the maximum is again identified from the updated noise loudness array, and one bit is assign to the corresponding
 30 frequency. This process continues until all available bits are exhausted.

For the 32 and 24 kb/s TPC coder, each of the 33 frequencies can receive bits during adaptive bit allocation. For the 16 kb/s TPC coder, on the other hand, better speech quality can be achieved if the coder assigns bits only to the frequency range of 0 to 4 kHz (*i.e.*, the first 16 FFT coefficients) and synthesizes the residual FFT coefficients in the higher frequency band of 4 to 8 kHz using the high-frequency synthesis processor 140.

Note that since the quantized LPC coefficients a are also available at the TPC decoder, there is no need to transmit the bit allocation information. This bit allocation information is determined by a replica of the hearing model quantizer control processor 50 in the decoder. Thus, the TPC decoder can locally duplicate the encoder's adaptive bit allocation operation to obtain such bit allocation information.

e. Quantization of Transform Coefficients

The transform coefficient quantizer 120 quantizes the transform coefficients contained in tc using the bit allocation signal ba . The DC term of the FFT is a real number, and it is scalar quantized if it ever receives any bit during bit allocation. The maximum number of bits it can receive is 4. For second through the 16th FFT coefficients, a conventional two-dimensional vector quantizer is used to quantize the real and imaginary parts jointly. The maximum number of bits for this 2-dimension VQ is 6 bits. For remaining FFT coefficients, a conventional 4-dimensional vector quantizer is used to jointly quantize the real and imaginary parts of two adjacent FFT coefficients. After the quantization of transform coefficients is done, the resulting VQ codebook index array IC contains the main information of the TPC encoder. This index array IC is provided to the multiplexer 180, where it is combined with side information bits. The result is the final bit-stream, which is transmitted through a communication channel to the TPC decoder.

The transform coefficient quantizer 120 also decodes the quantized values of the normalized transform coefficients. It then restores the original gain levels of these transform coefficients by multiplying each of these

coefficients by the corresponding elements of *mag* and the quantized linear gain of the corresponding frequency band. The result is the output vector *dtc*.

f. High-Frequency Synthesis and Noise Fill-In

- 5 For the 16 kb/s coder, adaptive bit allocation is restricted to the 0 to 4 kHz band, and processor 140 synthesizes the 4 to 8 kHz band. Before doing so, the hearing model quantizer control processor 130 first calculates the ratio between the LPC power spectrum and the masking threshold, or the signal-to-masking-threshold ratio (SMR), for the frequencies in the 4 to 7 kHz band.
- 10 The 17th through the 29th FFT coefficients (4 to 7 kHz) are synthesized using phases which are random and magnitude values that are controlled by the SMR. For those frequencies with $SMR > 5$ dB, the magnitude of the FFT coefficients is set to the quantized linear gain of the high-frequency band. For those frequencies with $SMR \leq 5$ dB, the magnitude is 2 dB below the
- 15 quantized linear gain of the high-frequency band. From the 30th through the 33rd FFT coefficients, the magnitude ramps down from 2 dB to 30 dB below the quantized linear gain of the high-frequency band, and the phase is again random.

For 32 and 24 kb/s coders, bit allocation is performed for the entire

20 frequency band as described. However, some frequencies in the 4 to 8 kHz band may still receive no bits. In this case, the high-frequency synthesis and noise fill-in procedure described above is applied only to those frequencies receiving no bits.

After applying such high-frequency synthesis and noise fill-in to the

25 vector *dtc*, the resulting output vector *qtc* contains the quantized version of the transform coefficients before normalization.

g. Inverse Transform and Filter Memory Updates

The inverse transform processor 150 performs the inverse FFT on the

30 64-element complex vector represented by the half-size 33-element vector

qtc. This results in an output vector *qtt*, which is the quantized version of *tt*, the time-domain target vector for transform coding.

With zero initial filter states (filter memory), the inverse shaping filter 160, which is an all-zero filter having *awc* as its coefficient array, filters the vector *qtt* to produce an output vector *et*. The adder 170 then adds *dh* to *et* to obtain the quantized LPC prediction residual *dt*. This *dt* vector is then used to update the internal storage buffer in the closed-loop pitch tap quantizer and pitch predictor 80. It is also used to excite the internal shaping filter inside the zero-input response processor 50 in order to establish the correct filter memory in preparation for the zero-input response generation for the next subframe.

C. An Illustrative Decoder Embodiment

An illustrative decoder embodiment of the present invention is shown in Figure 2. For each frame, the demultiplexer 200 separates all main and side information components from the received bit-stream. The main information, the transform coefficient index array *IC*, is provided to the transform coefficient decoder 235. In order to decode this main information, adaptive bit allocation must be performed to determine how many of the main information bits are associated with each quantized transform coefficient.

The first step in adaptive bit allocation is the generation of quantized LPC coefficients (upon which allocation depends). The demultiplexer 200 provides the seven LSP codebook indices *IL*(1) to *IL*(7) to the LPC parameter decoder 215, which performs table look-up from the 7 LSP VQ codebooks to obtain the 16 quantized LSP coefficients. The LPC parameter decoder 215 then performs the same sorting, interpolation, and LSP-to-LPC coefficient conversion operations as in blocks 345, 350, and 355 in Figure 3.

With LPC coefficient array *a* calculated, the hearing model quantizer control processor 220 determines the bit allocation (based on the quantized LPC parameters) for each FFT coefficient in the same way as processor 130

in the TPC encoder (Figure 1). Similarly, the shaping filter coefficient processor 225 and the shaping filter magnitude response processor 230 are also replicas of the corresponding processors 30 and 100, respectively, in the TPC encoder. Processor 230 produces *mag*, the magnitude response of the
5 shaping filter, for use by the transform coefficient decoder 235.

Once the bit allocation information is derived, the transform coefficient decoder 235 can then correctly decode the main information and obtain the quantized versions of the normalized transform coefficients. The decoder 235 also decodes the gains using the gain index array *IG*. For each subframe,
10 there are two gain indices (5 and 7 bits), which are decoded into the quantized log gain of the low-frequency band and the quantized versions of the level-adjusted log gains of the mid- and high-frequency log gains. The quantized low-frequency log gain is then added back to the quantized versions of the level-adjusted mid- and high-frequency log gains to obtain the
15 quantized log gains of the mid- and high-frequency bands. All three quantized log gains are then converted from the logarithmic (dB) domain back to the linear domain. Each of the three quantized linear gains is used to multiply the quantized versions of the normalized transform coefficients in the corresponding frequency band. Each of the resulting 33 gain-scaled,
20 quantized transform coefficients is then further multiplied by the corresponding element in shaping filter magnitude response array *mag*. After these two stages of scaling, the result is the decoded transform coefficient array *dtc*.

The high-frequency synthesis processor 240, inverse transform
25 processor 245, and the inverse shaping filter 250 are again exact replicas of the corresponding blocks (140, 150, and 160) in the TPC encoder. Together they perform high-frequency synthesis, noise fill-in, inverse transformation, and inverse shaping filtering to produce the quantized excitation vector *ef*.

The pitch decoder and interpolator 205 decodes the 8-bit pitch index *IP*
30 to get the pitch period for the last 3 subframes, and then interpolate the pitch period for the first two subframes in the same way as is done in the corresponding block 70 of the TPC encoder. The pitch tap decoder and pitch

predictor 210 decodes the pitch tap index IT for each subframe to get the three quantized pitch predictor taps b_{1k} , b_{2k} , and b_{3k} . It then uses the interpolated pitch period kpi to extract the same three vectors x_1 , x_2 , and x_3 as described in the encoder section. (These three vectors are respectively $kpi - 1$, kpi , and $kpi + 1$ samples earlier than the current frame of dt .) Next, it computes the pitch-predicted version of the LPC residual as

$$dh = \sum_{i=1}^3 b_{ik} x_i.$$

The adder 255 adds dh and et to get dt , the quantized version of the LPC prediction residual d . This dt vector is fed back to the pitch predictor inside block 210 to update its internal storage buffer for dt (the filter memory of the pitch predictor).

The long-term postfilter 260 is basically similar to the long-term postfilter used in the ITU-T G.728 standard 16 kb/s Low-Delay CELP coder.

The main difference is that it uses $\sum_{i=1}^3 b_{ik}$, the sum of the three quantized pitch taps, as the voicing indicator, and that the scaling factor for the long-term postfilter coefficient is 0.4 rather than 0.15 as in G.728. If this voicing indicator is less than 0.5, the postfiltering operation is skipped, and the output vector fdt is identical to the input vector dt . If this indicator is 0.5 or more, the postfiltering operation is carried out.

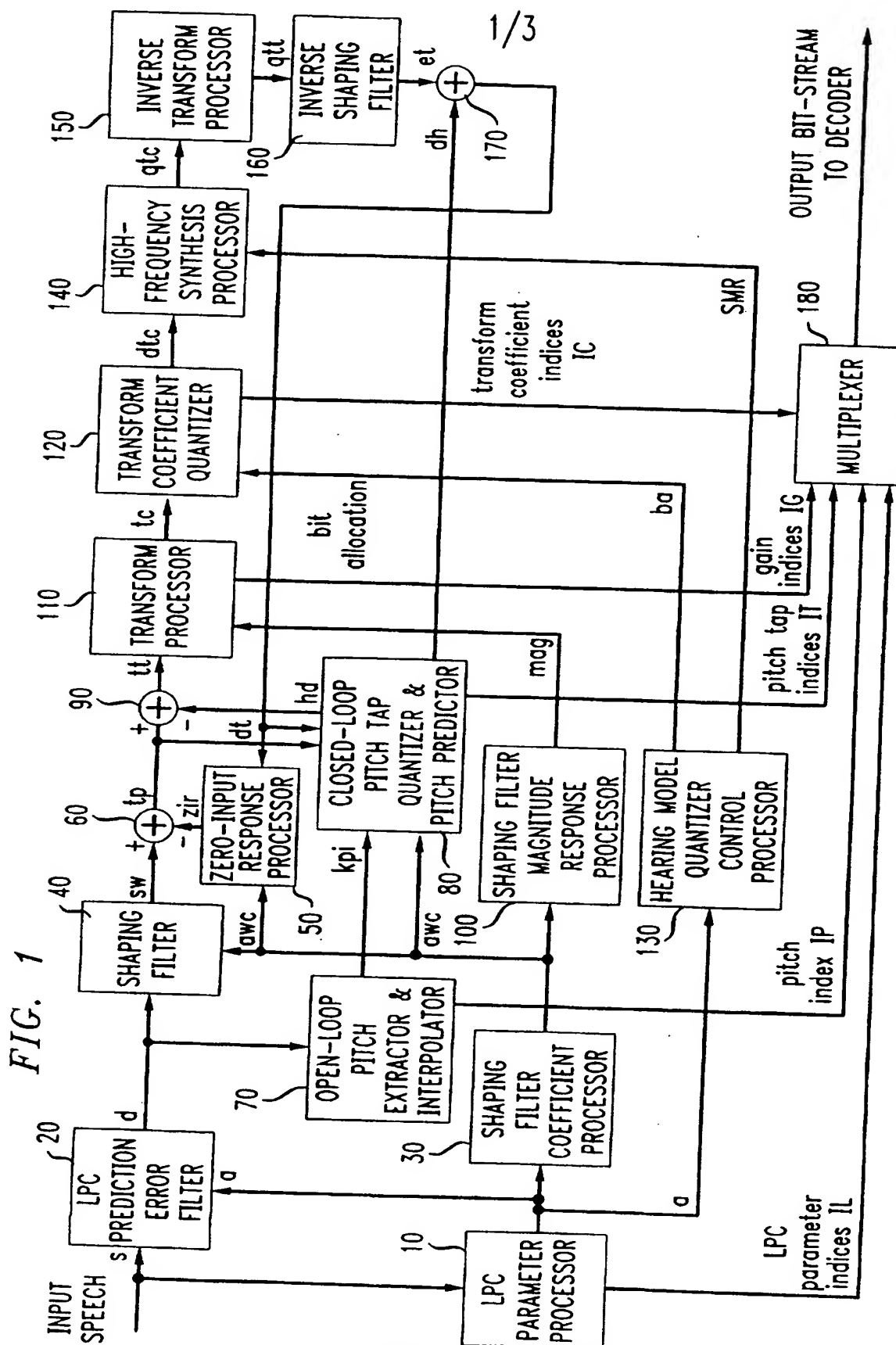
The LPC synthesis filter 265 is the standard LPC filter — an all-pole, direct-form filter with the quantized LPC coefficient array a . It filters the signal fdt and produces the long-term postfiltered, quantized speech vector st . This st vector is passed through the short-term postfilter 270 to produce the final TPC decoder output speech signal fst . Again, this short-term postfilter 270 is very similar to the short-term postfilter used in G.728. The only differences are the following. First, the pole-controlling factor, the zero-controlling factor, and the spectral-tilt controlling factor are 0.7, 0.55, and 0.4, respectively, rather than the corresponding values of 0.75, 0.65, and 0.15 in G.728. Second, the coefficient of the first-order spectral-tilt compensation filter is

linearly interpolated sample-by-sample between frames. This helps to avoid occasionally audible clicks due to discontinuity at frame boundaries.

The long-term and short-term postfilters have the effect of reducing the perceived level of coding noise in the output signal *fst*, thus enhancing the
5 speech quality.

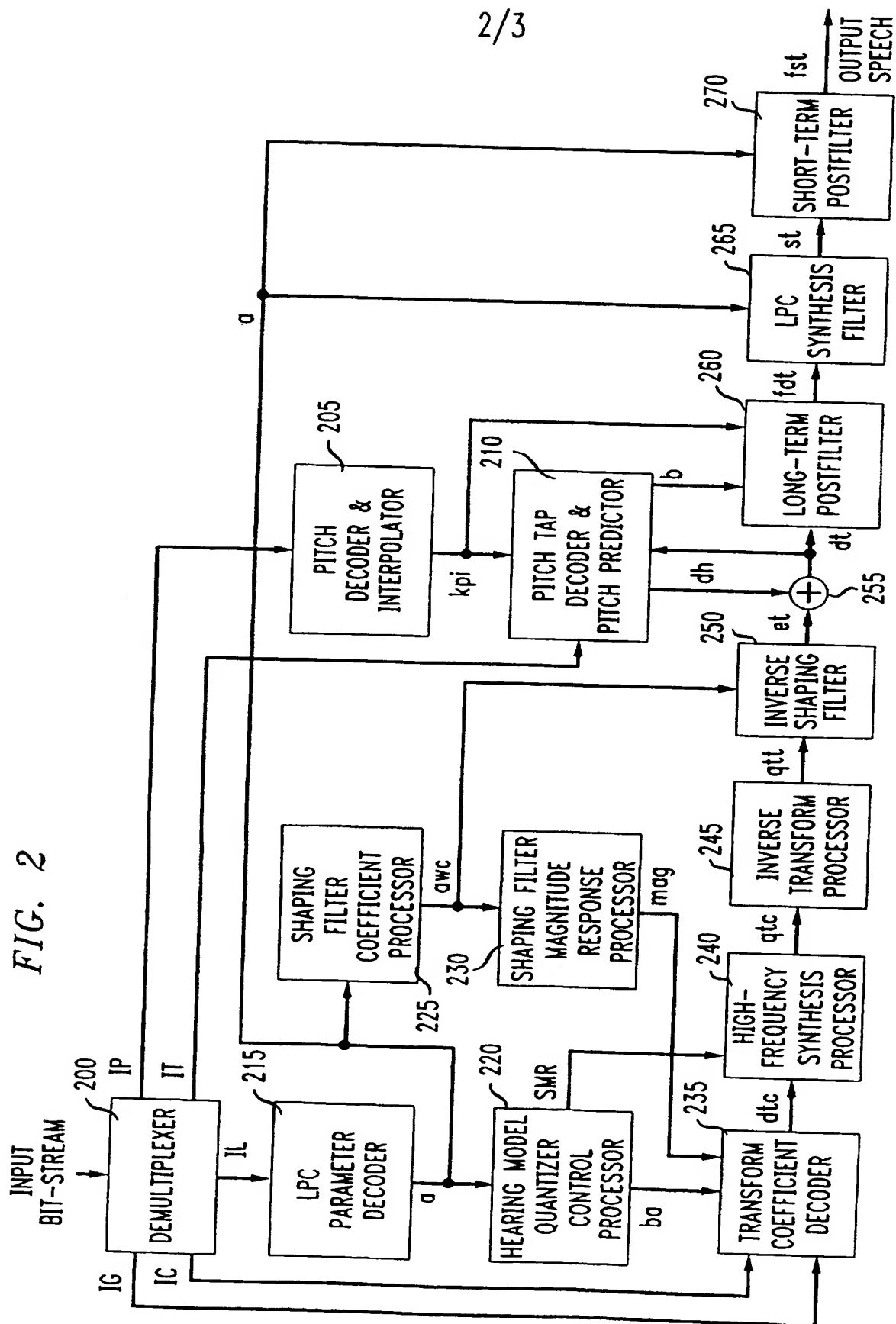
What is claimed is:

- 1 1. A method of coding a frame of a speech signal comprising the steps of:
 - 2 removing short-term correlations from the speech signal with use of a linear
 - 3 prediction filter to produce a prediction residual signal;
 - 4 determining an open-loop estimate of the pitch period of the speech signal
 - 5 based on the prediction residual signal;
 - 6 determining pitch filter tap weights for two or more subframes of the frame
 - 7 based on a quantized version of the prediction residual signal;
 - 8 forming a pitch prediction residual signal based on the open loop pitch period
 - 9 estimate, the pitch filter tap weights for the two or more subframes, and the
 - 10 prediction residual signal; and
 - 11 quantizing the pitch prediction residual signal.



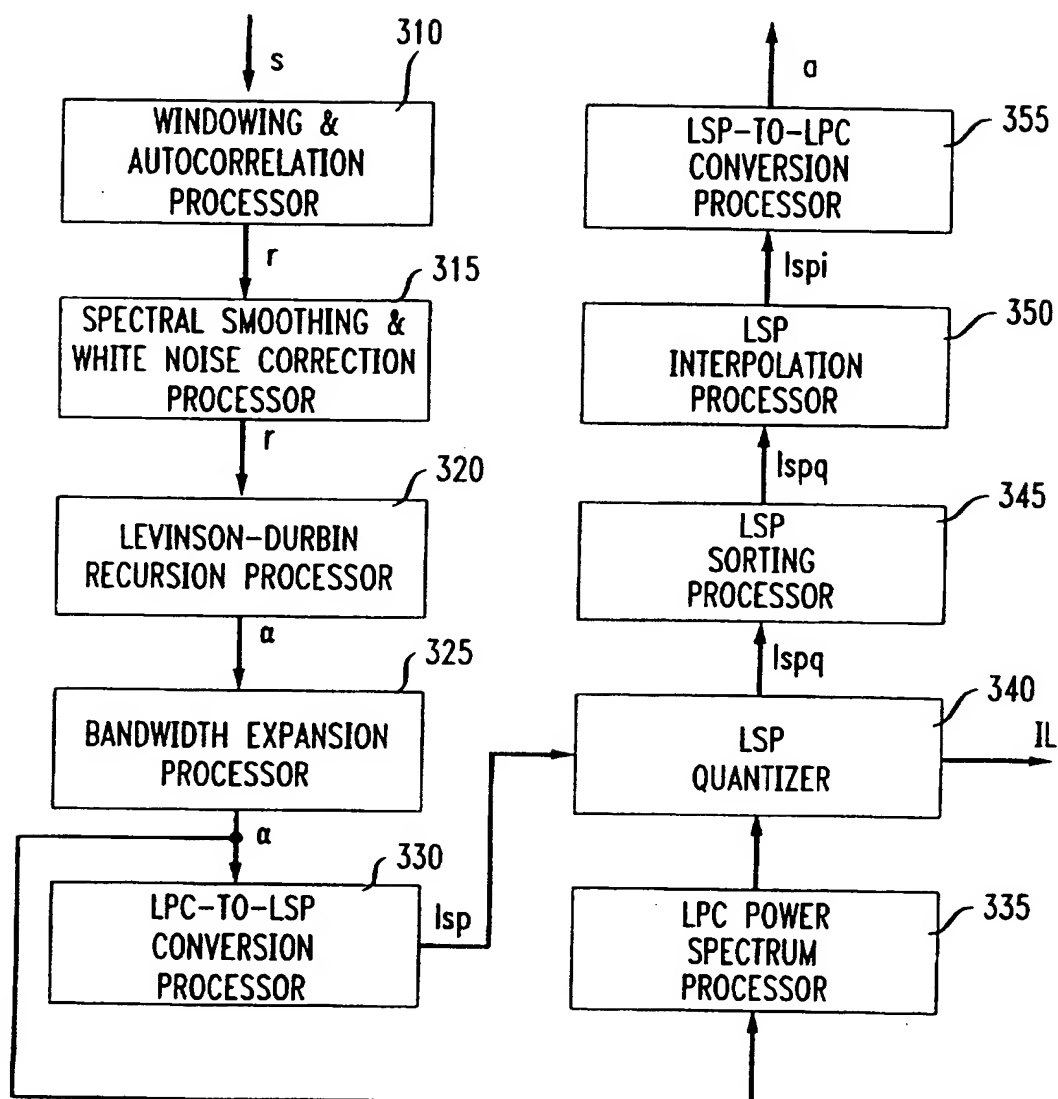
2/3

FIG. 2



3/3

FIG. 3



INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/02898

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : G10L 3/02, 9/00
US CL : 395/2.28, 2.1, 2.12
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 395/2.28, 2.1, 2.12

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
US CL 395/2.09-2.95


Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
APS; IEEE AND IEE CDROM DATABASE; SMART PATENT CDROM DATABASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	US 5,583,963 A (LOZACH) 10 December 1996 (10/12/96), column 4 line 64 - column 8 line 23, column 12 line 22-67.	1
X	US 5,012,517 A (WILSON ET AL) 30 April 1991 (30/04/91), column 7 line 12 - column 13 line 65.	1
Y	JOHNSON et al.; "Pitch-Orthogonal Code-Excited LPC.", Global Telecommunications Conference, IEEE GLOBECOM 90, pages 542-546.	1

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be part of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier document published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Z" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 28 APRIL 1997	Date of mailing of the international search report 06.06.1997
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-9508	Authorized officer  ALLEN MACDONALD Telephone No. (703) 305-9708

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/02898

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	OFER ET AL.; "A Unified Framework for LPC Excitation Representation in Residual Speech Coders"; Acoustics, Speech & Signal Processing Conference, IEEE ICASSP '89, pages 41-44 especially Section 5 page 44.	1
Y	DAVIDSON G. et al.; "Multiple-Stage Vector Excitation Coding of Speech Waveforms"; Acoustics, Speech & Signal Processing Conference, IEEE ICASSP '88, pages 163-166 especially Section 3 pages 164-166.	1

